

SABIO-RK, von Daten in der Publikation zur Suchlösung für Spezialisten

Wolfgang Müller¹  · Meik Bittkowski · Martin Golebiewski · Renate Kania · Maja Rey · Andreas Weidemann · Ulrike Wittig

Eingegangen: 28. Oktober 2016 / Angenommen: 22. Dezember 2016 / Online publiziert: 10. Februar 2017
© Der/die Autor(en) 2017. Dieser Artikel ist eine Open-Access-Publikation.

Zusammenfassung SABIO-RK ist eine Datenbank, in der Spezialisten aus der Systembiologie Daten aus biochemischen Publikationen suchen, finden, und in geeigneten Formaten extrahieren können. Der Artikel beschreibt, wie Kuratierung durch Experten, standardisierte Struktur, flexible Suche und einfacher Datenexport ineinandergreifen, um den Informationsbedarf der Nutzer zu befriedigen.

Schlüsselwörter SABIO-RK · Biocuration · Databases · Information Retrieval

1 Einleitung

Systembiologie [4] beschäftigt sich damit, Biologische Vorgänge als System zu interpretieren und zu verstehen. Experimentelle Daten werden gesammelt, integriert und in Modelle vereinigt, aus denen schließlich Erkenntnisse gewonnen, beziehungsweise neue Hypothesen generiert werden. Wichtig hierbei ist, dass es bei der Modellierung nicht um die *Datenmodellierung* geht, sondern um Modelle, die beschreiben, wie der stabile Zustand eines Systems zustandekommt, bzw. wie das biologische System auf Veränderungen antwortet.

Was den Systembiologischen Zyklus zu einer Herausforderung macht, ist, dass er die Grenzen zwischen Fachdisziplinen überschreitet. Experimentatoren arbeiten mit Bioinformatikern, aber auch Mathematikern, Biophysikern etc. zusammen. Natürlich gibt es auch innerhalb der Disziplinen etliche Spezialisierungen.

In der Systembiologie müssen also Daten über Fachgrenzen hinweg ausgetauscht und integriert werden. Eine wichtige Funktion hierbei haben wissenschaftliche Datenbank-Anwendungen, wie SABIO-RK [11]. SABIO-RK ist eine frei zugängliche, web basierte Datenbank (<http://sabiork.h-its.org/>). Ihr Fokus liegt auf kinetischen Daten für biochemische Reaktionen, die aus der Literatur extrahiert und in einem einheitlichen Format dargestellt werden. SABIO-RK wird von Modellierern, die mechanistische Modelle von biologischen Vorgängen erstellen, sowie von Experimentatoren verwendet.

In SABIO-RK greifen verschiedene Aspekte von Datenbanken und Information Retrieval ineinander. Es geht um Datenmodellierung, Datenkuratierung, Annotation, Indexierung und Anfragebearbeitung. Anders gesagt geht es darum: Wie werden Daten aufbereitet, wie werden sie mit anderen Daten vernetzt, wie werden sie schnell auffindbar und schnell zugreifbar gemacht?

2 Die Daten von SABIO-RK: Reaktionskinetische Daten

SABIO-RK bietet reaktionskinetische Daten: Für biochemische Reaktionen, die in oder zwischen Zellen ablaufen, wird die Reaktionsgleichung angegeben, und unter welchen Bedingungen die Reaktion wie schnell verläuft.

Diese Reaktionen sind Bestandteil von so genannten *Stoffwechselwegen*, also Reaktionsfolgen, in denen jeweils die Produkte einer Reaktion Eingang in eine nächste Reaktion finden. Abb. 1 gibt ein Beispiel.

Je nach Fachgebiet ist man an verschiedenen Sichtweisen auf den Stoffwechselweg interessiert.

Experimentatoren können SABIO-RK einsetzen, um ihre Messungen zu beurteilen. Ist die Messung im Einklang

✉ Wolfgang Müller
wolfgang.mueller@h-its.org

¹ HITS gGmbH, Heidelberg, Deutschland

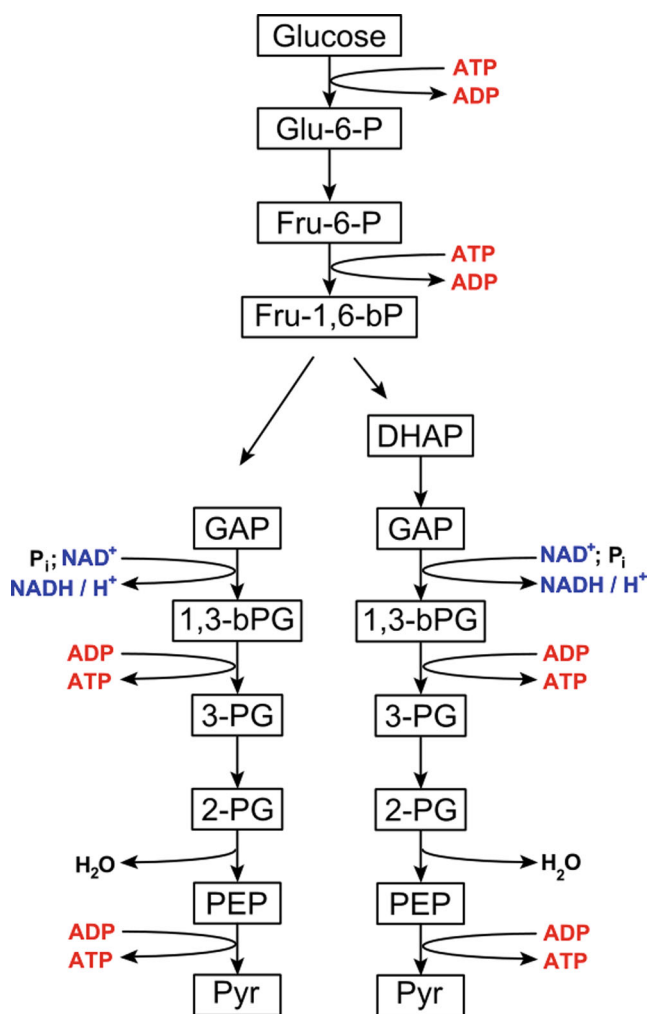


Abb. 1 Die Glykolyse ist ein Stoffwechselweg, d.h. eine Folge von Reaktionen, die in Zellen abläuft. In jeder Reaktion werden Substrate (also eingehende Moleküle) zu Produkten umgesetzt, beginnend mit der Umsetzung von Glukose zu Glukose-6-Phosphat (Glu-6-P), bei der Adenosintriphosphat eine Phosphatgruppe an die Glukose abgibt. Jeder Schritt wird durch ein Enzym katalysiert (nicht eingezeichnet), der erste Schritt durch die Hexokinase (EC 2.7.1.1). SABIO-RK speichert hierzu und zu anderen Stoffwechselwegen die reaktionskinetischen Daten. (Quelle der Grafik: de.wikipedia.org, gemeinfrei. Autor: Nutzer Ykrazuul)

mit Erwartungen aus der Literatur (ähnliche Reaktanten, ähnliche Bedingungen, ähnliche Kinetik), oder nicht?

Modellierer, die quantitative Modelle erstellen, werden daran interessiert sein, möglichst viele freie Parameter ihrer Modelle aus der Literatur zu bestimmen, und so die Zahl der zu „fittenden“, also durch Ausgleichsrechnung zu bestimmenden Parameter, zu minimieren und so zu belastbareren Modellen zu kommen.

3 Nutzer-Sicht

Die Geschwindigkeit einer Reaktion, die durch *Substrate*, *Produkte* und beteiligte *Enzyme* gegeben ist, hängt von verschiedenen Faktoren ab. Unter anderem:

- pH-Wert der Lösung,
- Konzentration der Reaktanten,
- Konzentration der Enzyme, die die Reaktion katalysieren.

In SABIO-RK ermöglichen wir unseren Nutzern, diese Information einfach und übersichtlich in einer Art „Datenblatt“ nachzulesen. Abb. 3 zeigt ein solches „Datenblatt“, das in SABIO *Entry* heißt.

Der Nutzer kennt die Reaktion, nicht aber die Parameter, d.h. es werden Lookup-Suchen durchgeführt. Sehr häufig wird aber in SABIO-RK auch die explorative Suche eingesetzt. Hier geht es darum zu erfahren, welche *möglichst ähnlichen* Reaktionen in SABIO-RK referenziert sind, zu welchen *möglichst ähnlichen* Organismen Daten vorhanden sind.

Die allermeisten Daten in SABIO-RK sind öffentlich zugänglich, da sie aus öffentlich zugänglichen Publikationen extrahiert sind. Jedoch ist es auch möglich, Daten zu nicht-publizierten Experimenten in SABIO-RK zu speichern. Zu einzelnen Datensätzen können flexibel Rechte an Nutzer und Nutzergruppen vergeben werden.

4 Kuratierung von Daten

Damit die Daten nach Parametern auffindbar sind, verwendet SABIO-RK *Ontologien* und *Standard-Identifikatoren*.

- **Ontologien** ermöglichen es hierbei, Relationen zwischen den Daten aufzudecken. Man kann beispielsweise über die Ontologie *NCBI Taxonomy* [6] herausfinden, dass *Saccharomyces cerevisiae* (Hefe), ein Pilz ist.
- **Standard-Identifikatoren** ermöglichen es, Entitäten eindeutig zu identifizieren. CHEBI:17234 bezeichnet beispielsweise das Molekül Glukose, für das die Datenbank ChEBI (Chemical Entities of Biological Interest) 5 Synonyme nennt.

All diese Daten müssen jedoch zunächst einmal eingepflegt werden. Die Quelle für derartige Daten ist immer noch die wissenschaftliche Publikation.

Wenn man analysiert, wo in einer Publikation Daten für SABIO-RK entries zu finden sind, stellt sich heraus, dass diese so weit über die Veröffentlichung verteilt sind, und gleichzeitig so stark inhaltlich zusammenhängen, dass für die Erstellung von SABIO-RK-Einträgen ein volles Verständnis der Veröffentlichung vonnöten ist. Man kann nicht im Vorhinein sagen, welche Daten wo in der Veröffent-

SABIO-RK
Biochemical Reaction Kinetics Database

Home Search Services Web Services News Documentation Publications Statistics Links About

Search

glucose AND Pathway: "glycolysis/gluconeogenesis" AND Enzymename: "hexokinase"

Advanced Search

OR Enzymename Add & Search

Filter Options

Enzyme

☒ Wildtype ☒ Mutant ☐ Recombinant

Kinetic Data

☐ Rate Equation

Reaction

☐ Transport Reaction

Environmental Conditions

pH: 0 - 14 Temperature: -10 C° - 115 C°

Source

☒ Direct Submission ☐ Entries inserted since: 14/10/2008

☒ Publication ☒ BioModel

Entry View Reaction View Visual Search (beta)

Total number of kinetic law entries found: **280** ☐ expand all displayed entries

1 2 3 4 5 6 7 8 9 10 .. 19 Next

Kinetic data	Reaction	Enzyme			Tissue	Organism	Parameter (besides concentration)	Environment		Add to export cart?
		ECNumber	Protein	Variant				°C	pH	

Abb. 2 Eine Anfrage, die nach dem Enzym Hexokinase und dem Stoffwechselweg Glykolyse sucht. Die oben als Text gegebene Anfrage wird mithilfe des Webformulars direkt unterhalb der Anfragezeile erstellt. Man sieht auch die Kopfzeile der Resultattabelle. Zur Glykolyse siehe Abb. 1

lichung zu finden sind. Zwar würde man dies beim definierten Aufbau von wissenschaftlichen Veröffentlichungen (Einleitung, Material/Methoden, Resultate, Diskussion), im Dies macht einen Einsatz von natürlichsprachlicher Datenverarbeitung zur Extraktion von Daten für SABIO-RK schwierig bis unmöglich [12].

Darum werden die Daten für SABIO-RK von Hand kuratiert.

Die Kuratierung in SABIO-RK findet in einem zweistufigen Prozess statt.

In einem ersten Schritt wird die eigentliche *Daten-Extraktion* durchgeführt. Hierzu liest eine wissenschaftliche Hilfskraft die Publikation und pflegt die Daten in eine Web-Eingabemaske ein.

Im zweiten Schritt werden diese Daten *kontrolliert, korrigiert, standardisiert und annotiert*. Dieser Schritt wird von den SABIO-RK Kuratorinnen durchgeführt. Die Kuratorinnen sind Biologinnen und Biochemikerinnen mit langjähriger Erfahrung, die miteinander in engem Kontakt stehen, um eine konsistente Aufbereitung der Daten zu gewährleisten.

Bei der Kontrolle wird geprüft, ob die Kuratorin die Entscheidungen der wissenschaftlichen Hilfskraft nachvollziehen kann. Eventuell werden die Daten korrigiert.

Die Standardisierung betrifft z.B. die Identifikatoren. Sind beispielsweise gleiche Reaktanten in SABIO-RK gleich benannt?

Die Annotation betrifft die *Vernetzung* der Daten innerhalb von SABIO-RK mit Daten außerhalb von SABIO-RK. Beispielsweise werden Informationen über Proteine in der UniProt-Datenbank verlinkt [13], Informationen über Reaktanten mit ChEBI [1], weitere Informationen zu Enzymen mit BRENDA [7]. Diese Datenbanken enthalten jeweils komplementäre Daten zu den entsprechenden Entitäten in SABIO-RK.

5 Ein Überblick über Suche in SABIO-RK

Es gibt verschiedene Arten der Suche in SABIO-RK:

1. Die Suche durch die Freitext-Suchleiste. Ein eingegebener Suchterm wird auf alle Bereiche von SABIO-RK angewandt. Beispiel: *Liver*
2. Die Expertensuche: Ein eingegebener Suchterm wird auf einen Teilbereich der Daten von SABIO-RK eingeschränkt. Beispielsweise kann man nach einem Organismus suchen. Bei der Suche mit *Organism: saccharo-*

Abb. 3 Ein Resultat der Anfrage aus Abb. 2. Ein SABIO-RK Entry, der Informationen über Hexokinase liefert

Entry ID: 44921

General information							
Organism	Asterias amurensis						
Tissue	pyloric cecum ↗						
EC Class	2.7.1.1						
SABIO reaction id	793						
Variant	wildtype						
Experiment Type	in vitro						
Pathways	Glycolysis classical Glycolysis/Gluconeogenesis Starch and Sucrose metabolism						
Event Description	-						
Substrates							
name	location		comment				
D-Glucose	-		-				
ATP	-		-				
Products							
name	location		comment				
D-Glucose 6-phosphate	-		-				
ADP	-		-				
Modifiers							
name	location	effect	comment	protein complex			
hexokinase(Enzyme)	-	Modifier-Catalyst	purified	Hexokinase;			
Enzyme (protein data)							
	UniProtKB_AC	name	mol. weight (kDa)	deviation (kDa)			
subunit	-	-	-	-			
complex	-	-	49.0	1.0			
Kinetic Law							
type		formula	annotation				
Mixed product inhibition		-	SBO:0000265 ↗				
Parameter							
name	type	species	start val.	end val.	deviat.	unit	comment
B	concentration ↗	ATP	2.0	-	-	mM	-
A	concentration ↗	D-Glucose	0.05	0.1	-	mM	-
I	concentration ↗	D-Glucose 6-phosphate	0.0	0.24	-	mM	-
K _i	K _i ↗	D-Glucose 6-phosphate	0.09	-	-	mM	versus D-Glucose
K _i	K _i ↗	D-Glucose 6-phosphate	0.13	-	-	mM	versus D-Glucose
Experimental conditions							
	start value		end value		unit		
pH	7.4				-		
temperature	21.0				-		
buffer	100 mM Tris, 16 mM MgCl2, 1 mM Phosphoenolpyruvate, 0.5 mM NADH, 0.5 units/ml Pyruvate kinase, 0.5 units/ml Lactate dehydrogenase						
General comment							
Reference							
title	author	year	journal	volume	pages	PubMed	
Purification and properties of hexokinase from the starfish, Asterias amurensis	Mochizuki Y, Hori SH	1977	Biochem	81	1849-55	893376 ↗	

- myces* wird der Suchbegriff *Saccharomyces* nur auf das Feld „Organism“ angewandt. Beispiel: Abb. 5
- Die visuelle Suche: In Histogrammen und Tortendiagrammen sind die einzelnen Felder clickbar. Hiermit kann die Menge der Suchresultate weiter eingeschränkt werden. Ein Beispiel findet sich in Abb. 6.
 - Verweis auf andere Datenquellen: Für Organismen, Gewebe, Reaktanten, Proteine, Enzyme etc. gibt es jeweils spezielle Datenquellen, die aus SABIO-RK entsprechend referenziert sind und auf die SABIO-RK direkt verweist. Ebenso verweisen andere Datenquellen auf SABIO-RK. Es ist also eine Suche denkbar, die einen auf die UniProt-Datenbank führt, dann nach BRENDA und wieder zurück nach SABIO-RK.

Typische Beispiele für Suchen sind:

- Suche nach Daten zu Reaktionen in der Leber der Maus (*Mus musculus*), filtern der Daten nach *Wild type*, also unveränderten Proteinen.
- Blutzellen in Säugetieren.
- Reaktionen, die bei 37 °C und pH 6–8 gemessen wurden (also bei menschlicher Körpertemperatur und lebensnahem pH-Wert)
- Reaktionen mit Hexokinase (Abb. 1) in *Mus musculus*. Das Enzym hat die EC-Nummer 2.7.1.1. Detaillierte Informationen über Proteine, die z. B. in Maus diese Hexokinase Aktivität aufweisen, können über einen direkten Link zur UniProt Datenbank [13], erhalten werden.

Edit entry

Publication ID: 4970 Entry ID: 64914

Pathway		pathway		0			
Reaction		H ₂ O + 2-Lysophosphatidylcholine = 2-Lysophosphatidate + Choline		13349	Transport: <input type="checkbox"/> Reverse reaction		
Compounds							
Stoich.	Name	Name Abbr./Synonym Name	Role	Cell. location	Loc. ID	Complex Protein	Comment
						Protein Identifier	Protein Name
1	Enzyme		Modifier-Catalyst	extracellular	2		
1	H ₂ O		Substrate	extracellular	2		
1	2-Lysophosphatidylcholine		Substrate	extracellular	2		LPC (16:0)
1	2-Lysophosphatidate		Product	extracellular	2		LPA (16:0)
1	Choline		Product	extracellular	2		

Enter compound:

Enter compound(s):

Choose location:

Choose pathway:

Signalling event (reaction_based)

Abb. 4 Das SABIO-RK Input Interface. Hier werden die Daten aus Publikationen eingepflegt

Die meisten Nutzer sehen in SABIO-RK die Daten nicht nur an, sie exportieren sie, um sie in Modellen zu verwenden. Hierzu können die Daten markiert werden. Sie werden dann in einem „Einkaufswagen“ abgelegt, aus dem heraus sie dann als SBML-Dateien (Systems Biology Markup Language zum Austausch von Modellen) oder auch flexibel als Excel-Tabellen gespeichert werden können.

Darüber hinaus bietet SABIO-RK Web Services, die ebenfalls Daten als SBML exportieren. Auf Basis von diesen Webservices wurde z.B. eine Cytoscape-App [5] entwickelt, die SABIO-RK-Daten in der populären Visualisierungs-Anwendung Cytoscape [8] darstellt.

6 Die technische Seite von SABIO-RK

Technisch gesehen war SABIO-RK zunächst eine auf proprietären relationalen Datenbanken aufsetzende JDBC-Anwendung, die diese Daten als JSP (Java Server Pages) anzeigt. Inzwischen sind die meisten Teile von SABIO-RK als Grails/Groovy/GSP-Kombination auf Basis von PostgreSQL realisiert.

Da sich jeder Eintrag aus Informationen mehrerer Datenbanktabellen zusammensetzt und die verschiedenen Relationen zwischen einzelnen Objekten abgebildet werden müssen, ist das SABIO-RK Datenbankschema sehr kom-

plex. Zur Beantwortung einer Nutzeranfrage müssen an die Datenbank viele Anfragen gestellt werden. Ein Großteil dieser Anfragen ist schwer zu optimieren.

Weitere Last wird der Datenbank durch Access Control Lists aufgebürdet. Für jeden Eintrag muss zusätzlich bestimmt werden, ob er durch einen gegebenen Nutzer überhaupt angesehen werden darf.

Der Schlüssel zu einer einfachen Optimierung von SABIO-RK ist die Beobachtung, dass obgleich die *Entries* von SABIO-RK komplex aufgebaut sind, sie im Sinne eines Zugriffs atomar sind: Eine Person darf einen *Entry* sehen, oder sie darf den *Entry* nicht sehen. Der *Entry* wird als Ganzes angezeigt, oder er wird nicht angezeigt.

Damit sind die Voraussetzungen erfüllt, *Entries* als Ganzes zu cachen. Aus Geschwindigkeitsgründen speichert SABIO-RK ganze *Entries* als serialisierte JAVA-Objekte, die wir dann sehr einfach wieder deserialisieren können. Eine übliche SABIO-RK-Instanz hält sämtliche SABIO-RK *Entries* im RAM.

Damit ändert die Datenbank ihre Rolle. Ihre wichtigste Rolle ist die des Zwischenspeichers für die Sammlung der Daten, und als *Spezifikationsplattform* für die korrekte Struktur der Daten. In der Datenbank können nämlich einfach Constraints deklariert werden, die beispielsweise in einem Speicher für semistrukturierte Dokumente sehr viel schwerer aufrechtzuerhalten wären.

Abb. 5 Suchleiste sowie Expertensuche mit Auto-Vervollständigung. Die Nutzer haben die Wahl, entweder die Freitextsuche zu verwenden (die Suchterme werden auf alle Felder angewandt) oder Suchterme auf bestimmte Felder anzuwenden, z. B. Organismus. Bei Organismen ist auch die Suche nach ähnlichen Organismen mithilfe von Ontologien möglich. Erkennbar ist auf dieser Abbildung auch die Vorberechnung der Resultatgröße

The screenshot shows the SABIO-RK search interface. At the top, there is a main search bar containing the text 'blood'. To its right are buttons for 'Reset' and a help icon. Below this is an 'Advanced Search' section. It features a dropdown menu currently set to 'Organism', followed by a text input field containing 'sac'. To the right of this input is an 'Add & Search' button. A dropdown menu is open below the 'sac' input, displaying a list of suggestions with their respective result counts in parentheses: 'saccharomycotina (NCBI) (5)', 'saccharomyces (NCBI) (5)', 'saccharum officinarum (0)', 'saccharum hybrid (0)', and 'saccharum (NCBI) (0)'. At the bottom of the interface, there are three tabs: 'Entry View' (which is active), 'Reaction View', and 'Visual Search (beta)'.

Das Caching von *Entries* hat sehr positive Auswirkungen auf die Anfragebearbeitung, da die Datenbank nur noch für die Auswahl von *Entries* zuständig ist, nicht jedoch für deren Erstellung und Anzeige.

Die eigentliche Anfragebearbeitung wurde dadurch optimiert, dass wir SOLR [10] einsetzen. Die Auswahl der anzuzeigenden *Entries* ist üblicherweise eine Konjunktion von Auswahlkriterien, die sich auf eine Anfrage auf invertierten Listen mit anschließendem Schnitt der Resultatmengen abbilden lassen. Elegant lassen sich auch Sichtbarkeitsberechnungen auf diese Weise durchführen: Die gleiche Anfrage, die *Entries* nach Suchkriterien auswählt, kann auch auswählen, ob die gefundenen *Entries* dem Nutzer oder der Nutzerin gezeigt werden dürfen. Hierzu erstellen wir für jeden Nutzer eine Liste der Objekte, die er sehen darf, die ebenso wie die anderen Parameter als invertierte Liste gespeichert wird.

Im Ganzen haben uns die oben beschriebenen Optimierungen gegenüber einer Datenbank-Lösung eine Geschwindigkeits-Steigerung von Faktor 50 gebracht [9].

7 Optimierung als Basis für Autocompletion

Dieser Faktor 50 ist nicht nur eine theoretische Zahl, ermöglicht neue Funktionalität wie Auto-Completion. Die komplexen biologischen Begriffe sind zwar den Nutzern von SABIO-RK gut bekannt, bei längeren Wörtern sind Rechtschreibfehler jedoch immer möglich. Diese werden durch Autocompletion vermieden. Gleichzeitig gibt SABIO-RK eine *Vorschau*, wie viele Resultate für gewisse Anfrageterme zu erwarten sind. Häufig reicht dem Nutzer schon das Schreiben einiger weniger Buchstaben, um den richtigen

Anfrageterm zu sehen, wieviele Resultate eine Anfrage haben wird, beziehungsweise um zu sehen, dass etwas *nicht* in der Datenbank ist.

Diese Vorschau ist nur möglich, wenn man sehr schnell berechnen kann, wie viele *Entries* das Resultat einer gegebenen Anfrage sind. Die beschriebene Architektur mit SOLR findet dies hinreichend schnell heraus, um für mehrere Anfragen entsprechende Resultatgrößen zu bestimmen. Das Tippen wird von der Vorschau nicht unterbrochen.

Ebenso ist es mit dieser Architektur möglich, on-the-fly Statistiken der Resultate zu erstellen, und somit einen Überblick zu erstellen, der die Entscheidung über die Art der weiteren Suche beeinflusst. Wie viele Daten gibt es zu gegebenen Organismen, welche Daten liegen zu Reaktionen vor, die von einem gegebenen Enzym beeinflusst werden?

8 Erfahrungen vs. Evaluation

SABIO-RK enthält ca. 50.000 handkuratierte Einträge.

Über die Jahre haben wir die *Suche* von SABIO-RK in zwei Richtungen optimiert:

- Schnell das Gesuchte finden (schneller Erfolg)
- Schnell finden, dass es das Gesuchte nicht gibt (schnelles Scheitern)

Diese Faktoren lassen sich schnell auszählen und messen: Beim jeweiligen Vergleich haben wir gezählt, wie viele Clicks wir für ein Resultat brauchten, bzw. in welcher Zeit zu einem Resultat zu kommen war. Diese simplen Betrachtungen haben uns bei der Entwicklung geleitet.

Von SABIO-RK gibt es keine formelle Evaluation, sondern einen sehr großen A/B-Test: Die größte Umstellung

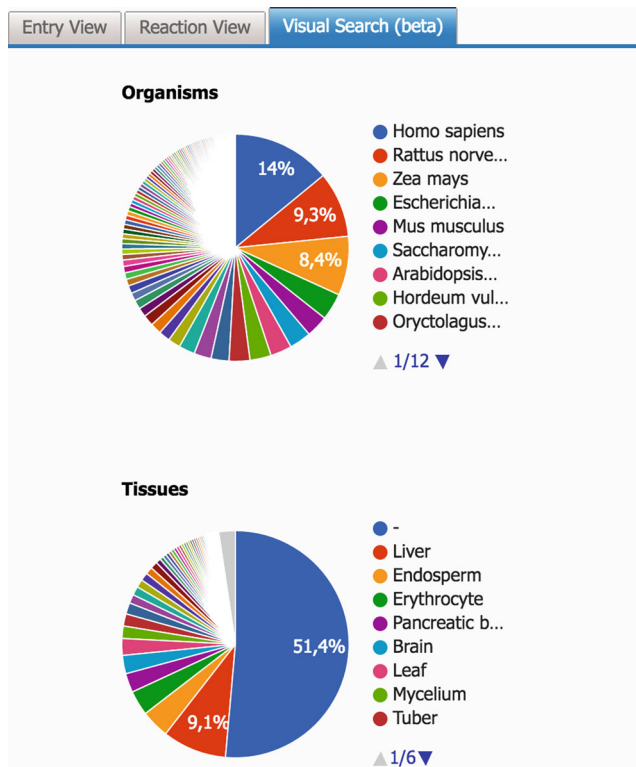


Abb. 6 Einfache Visualisierung der Verteilung eines Anfrageresultats auf Organismen und Gewebetypen. Die Sektoren des Tortendiagramms sind jeweils clickbar. Ein Click führt zu einer weiter verfeinerten Anfrage

des Nutzer Interfaces haben wir parallel zur alten Version eingeführt. Das Faktum, dass nach kurzer Zeit keiner mehr die alte Version nutzte, überzeugte uns von der Überlegenheit des neuen Interfaces.

Die meisten kurzfristigeren Informationen über das Nutzerverhalten bekommen wir durch (i) direkte Beobachtung bei Workshops und mit kleinen Mengen von Testern, sowie durch (ii) Nutzungszahlen.

Die Suche von SABIO-RK wird mit Funktionalität komplettiert, die es den Nutzern erleichtern soll, die Suchergebnisse weiterzuverwenden. Dies umfasst Export-Funktionalität und Webservices. Webservices ermöglichen weitere Funktionalität, die von anderen Anbietern kommen, z. B. die bereits beschriebene Cytoscape-App, sowie z. B. das Modellierungstool COPASI, das seinen Nutzern ermöglicht, SABIO-RK-Daten zu importieren, ohne das Werkzeug zu verlassen [2, 5].

Der Kern des Erfolges von SABIO-RK liegt jedoch in den Daten, d. h. der eigentlichen Strukturierung der Daten sowie der Vernetzung mit anderen Datenquellen. Die Strukturierung ermöglicht eine leichte Wiederverwendung der Daten, sowie einen guten Überblick über die Daten.

9 Zusammenfassung und Ausblick

Wir haben die Datenbank SABIO-RK vorgestellt, die reaktionskinetische Daten auf einfache Art flexibel suchbar macht und zum Export bereitstellt. Wir haben einen Eindruck von den Suchmöglichkeiten vermittelt und gezeigt, welche Optimierungen diese ermöglicht haben.

Schließlich sind wir kurz auf unsere Erfahrungen mit Nutzern eingegangen.

Die Erweiterungsmöglichkeiten sind vielfältig. Offensichtlich ist Kuratierung ein wichtiges Thema. Wie kann man schneller, kostengünstiger, besser kuratieren und so für mehr Quantität bei gleichzeitig noch höherer Qualität der Daten sorgen? Ist eine Integration von teilautomatischen Methoden bei diesen komplexen Daten möglich?

Auf der anderen Seite versuchen wir, den Daten bereits innerhalb von SABIO-RK mehr Wert zu geben, den Überblick über die Daten zu verbessern. Hierfür sind neben Detailverbesserungen größere Erweiterungen in Entwicklung, die die visuelle Suche innerhalb von SABIO-RK weiter verbessern werden.

Acknowledgements SABIO-RK verdankt seine inzwischen zehnjährige Existenz dem HITS, der Klaus Tschira Stiftung, dem BMBF (HepatoSys II (FKZ 0313078C), SysMO-LAB (FKZ 03 16 182E), SBEPo (FKZ 0316182E), Virtual Liver Network (FKZ 0315749), de.NBI(FKZ 0315749)), sowie der DFG (Projekt Integrierte Immunoblot-Umgebung). SABIO-RK hat viel von Zusammenarbeit mit Nutzern profitiert, ganz besonders ist zu nennen Ursula Kummer, die als (zunächst zukünftige) Nutzerin von Anfang an SABIO-RK beeinflusst und begleitet hat. SABIO RK wurde initial unter Leitung von Isabel Rojas (1968–2013) konzipiert, der wir dankbar und in Freundschaft gedenken.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Literatur

- Hastings J, de Matos P, Dekker A, Ennis M, Harsha B, Kale N, Muthukrishnan V, Owen G, Turner S, Williams M, Steinbeck C (2013) The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic Acids Res* 41:D456–463. doi:10.1093/nar/gks1146
- Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N, Singhal M, Xu L, Mendes P, Kummer U (2006) COPASI: a COMplex PATHway Simulator. *Bioinformatics* 22:3067–3074
- Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar Autumn A, Dronov Sergey GED, Ginkel M, Gor V, Goryanin II, Hedley WJ, Hodgman TC, Hofmeyr J-HS, Hunter PJ, Juty NS, Kasberger JL, Kremling A, Kummer U, Le Novère N, Loew LMLD, Mendes P, Minch E, Mjolsness ED, Nakayama Yoichi NMR, Nielsen Poul F, Sakurada T, Schaff JC, Shapiro BE, Shimizu TS, Spence

- HD, Stelling J, Takahashi K, Tomita M, Wagner JM, Wang J et al (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 9(4):524–531
4. Hiroaki K (2002) Systems Biology: A Brief Overview. *Science* 295(5560):1662–1664. doi:[10.1126/science.1069492](https://doi.org/10.1126/science.1069492)
 5. König M (2016) cy3sabiork: A Cytoscape app for visualizing kinetic data from SABIO-RK [version 1; referees: 2 approved with reservations doi:[10.12688/f1000research.9211.1](https://doi.org/10.12688/f1000research.9211.1) (F1000Research 5:1736)]
 6. Sayers EW, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio MRE, Federhen S, Feolo M, Geer LY, Helmberg W, Kapustin Y, Landsman D, Lipman DJ, Madden TL, Maglott DR, Miller V, Mizrahi I, Ostell J, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Souvorov A, Starchenko G, Tatusova TA, Wagner L, Yaschenko E, Jian Y (2008) Database resources of the National Center for Biotechnology Information. *Nucl Acids Res* 37 (suppl_1):D5–D15. doi:[10.1093/nar/gkn741](https://doi.org/10.1093/nar/gkn741)
 7. Schomburg I, Hofmann O, Baensch C, Chang A, Schomburg D (2000) Enzyme data and metabolic information: BRENDA, a resource for research in biology, biochemistry, and medicine, S 109–118 (*Gene Funct. Dis.* 3–4)
 8. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11):2498–2504
 9. Shi L, Jong L, Müller W, Algaa E (2011) 50x Faster: Speeding up an SQL-based legacy system with few changes. *Lecture Notes in Informatics (LNI) – Proceedings, Series of the Gesellschaft für Informatik (GI)*, Bd. P-192.
 10. <http://lucene.apache.org/solr/>. Zugegriffen: 16.1.2017
 11. Wittig U, Kania R, Golebiewski M, Rey M, Shi L, Jong L, Algaa E, Weidemann A, Sauer-Danzwith H, Mir S, Krebs O, Bittkowski M, Wetsch E, Rojas I, Müller W (2012) SABIO-RK – database for biochemical reaction kinetics. *Nucleic Acids Res* 40(D1):D790–D796
 12. Wittig U, Kania R, Bittkowski M, Wetsch E, Shi L, Jong L, Golebiewski M, Rey M, Weidemann A, Rojas I, Müller W (2014) Data extraction for the reaction kinetics database SABIO-RK. *Perspect Sci* 1:33–40
 13. The UniProt Consortium (2015) UniProt: a hub for protein information. *Nucleic Acids Res* 43:D204–D212